Audio classification of Urbansound8k dataset using machine learning and deep learning

Sukkur IBA University

Presented by MSCS21 Misbah Aziz and Bilal Ahmed

Keywords: sound identification, voice classification, audio analysis, ML, DL, urbansound8k dataset.

Abstract: In each part of human existence, sound assumes a significant part. From individual security to basic observation, sound is a vital component to foster the mechanized frameworks for these fields, Sound recognizable proof is considered as serious deal in sound analysis and sound classification field, so in this project we have utilized machine learning models and deep learning algorithms, for example, DNN for the ID of various sounds from various voices datasets with the assistance of Different classifier to which class it has a place as indicated by classes, for exploratory data analysis we have taken urbansound8k dataset which contains 8732 named passages which contains 10 unique labelled classes the voice of siren, children playing, street music, dog bark, car horn, air conditioner, drilling, jackhammer, gunshot and engine idling.

Introduction: The classification of ecological sound has loads of applications in enormous scope and content based mixed media ordering and recovery. These days, in view of mixed media sensor organizations and enormous amounts of online sight and sound substance, individuals focus harder on sound classification in metropolitan conditions. There are different situations and fields, for example, wellbeing administrations and correspondence where a precise metropolitan sound acknowledgment framework can help individuals [1].

Living in a world encompassed by various types of sound from various sources, our cerebrum and hear-able framework is continually recognizing each sound that it hears, in its direction [1]. Grouping audio or sound has been a significant field of examination for a long time now and there have been many attempted and tried techniques with various models and features which have demonstrated to be helpful and exact. Classification of audio can go from fields like media, bioacoustics checking, and gatecrasher identification in untamed life regions to audio observation, and ecological sounds [2].

In the utilization of artificial insight and progressed sound advancements in creature sound classification , certain difficulties are as yet confronted, for example, the disturbances of foundation commotion [3].

Man-made consciousness innovations are progressively being applied to issues in natural exploration and protection. In the field of untamed life nature, the utilization of man-made consciousness in blend with late advances in review procedures has empowered scientists to gather data on

species events at a lot more extensive spatial and worldly scales than were beforehand conceivable [4].

As of late, deep neural organizations (DNNs) and all the more specifically Convolution neural organizations (CNNs) have been utilized to consequently gain highlight repre-sentations from complex data [5].

Sound acknowledgment issue

Comprises of three distinct stages as pre-processing of signals,

Extraction of explicit features and their classification. Signal

Pre-processing separates the info sign to various fragments

This utilized for extracting related features. Include extraction

Diminishes the size of data and represent the mind boggling data as

Include vectors [6].

Sounds contain rich information and assist with peopling sense the environments around them. Individuals can recognize complex sounds and filter out the meaningful information. Along these lines, noise is dropped and the useful information is refined. Today, sensors can undoubtedly gather tons of useful audio data; in any case, processing them to get meaningful information remains burdensome. Many analysts desire to design a human-like machine to reduce this kind of issues [7].

Thus, The fundamental objective of our project is to:

1. Extracting audio features from urbansound8k dataset.

2. Classifying the category of sound based on class with numeric identifier.

In this project, by utilizing UrbanSound8K dataset, we prepared models including ANN (Artificial neural networks), DT (decision tree), RF (random forest) and SVM (simple vector machine to order sounds in metropolitan conditions.

There are three distinct stages that are connected to the

Classification of sound; pre-processing of the sound sign,

Explicit otherworldly element extraction, lastly the classification of the sound sign. Signal pre-processing samples the input sound sign into different pieces which are used for removing fundamental features.
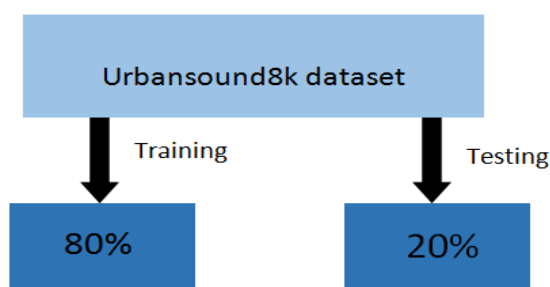
Methodology:

Dataset:

Figure 1

- We have done EDA using urban sound dataset which contains 7832 features.
- Urbansound8k dataset available on link:

  https://urbansounddataset.weebly.com/download-urbansound8k.html
- Urbansound dataset uses 10 different classes with different labels, Each audio document has a name formatted as [fsID]-[classID]-

[occurrenceID]-[sliceID].wav, where

• fsID is the Free sound ID of the recording from where this extract (slice) is taken.

• classID is a numeric identifier of the sound class.

• occurrenceID is a numeric identifier to distinguish unique.

events of the sound inside the original recording.

• sliceID is a numeric identifier to distinguish various cuts taken from the same event.

Pre-processing:

Given is the overview of classes with specified numeric identifier:

Table 1:

Category (numeric identifier)	Class
0	Air conditioner
1	Car horn
2	Children playing
3	Dog bark
4	drilling
5	Engine idling
6	Gunshot
7	jackhammer
8	siren
9	Street music

Software: We have used python 3 with Jupiter notebook to implement the audio classification code.

Features : We have used Librosa library to extract the MFCC features which used two types of signals one is mono and other one is stereo we have used mono type for

audio analysis in this given dataset and the length of features is 40 used in our project.

Models and input:

We have used four different models to train and test our data following is the short description:

Table 2:

Model name	Input
ANN	We can foster a solitary MLP neural network model that can make both regression and classification forecasts for a solitary input. This is known as a multi-yield model and can be created utilizing the Keras API.
DT	A decision tree is a graphical portrayal of all potential answers for a decision in light of specific circumstances. On each progression or node of a decision tree, utilized for classification analysis purpose.
RF	A random forest utilized to take care of regression and classification issues. It uses group learning, which is a method that joins numerous classifiers to give answers for complex issues.
SVM	The upsides of support vector machines are: Effective in high layered spaces. Still powerful in situations where number of aspects is more noteworthy than the quantity of samples.

Results and discussion:

✓ Model training and accuracy:
Given is the Accuracy table of models we have used in our project

Table 3:

Model	Accuracy
ANN	75.8%
DT	69.8%
RF	88.3%
SVM	61.1%

✓ ANN Model:

Table 4:

```
Model: "sequential"

Layer (type)                    Output Shape              Param #
=================================================================
dense (Dense)                   (None, 100)               4100

activation (Activation)         (None, 100)               0

dropout (Dropout)               (None, 100)               0

dense_1 (Dense)                 (None, 200)               20200

activation_1 (Activation)       (None, 200)               0

dropout_1 (Dropout)             (None, 200)               0

dense_2 (Dense)                 (None, 100)               20100

activation_2 (Activation)       (None, 100)               0

dropout_2 (Dropout)             (None, 100)               0

dense_3 (Dense)                 (None, 10)                1010

activation_3 (Activation)       (None, 10)                0
=================================================================
Total params: 45,410
Trainable params: 45,410
Non-trainable params: 0
```

✓ Classes and its samples distribution

Dog bark	1000
Children playing	1000
Air conditioner	1000
Street music	1000
Engine idling	1000
Jackhammer	1000
Drilling	1000
Siren	929
Car horn	429
Gunshot	374

✓ Sample of audio signal generated using Librosa:
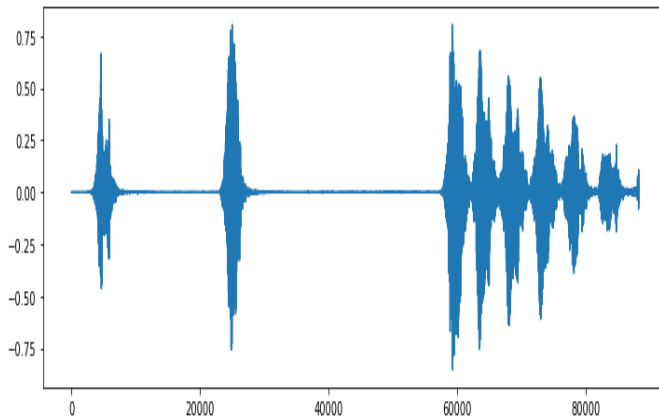
Figure 2
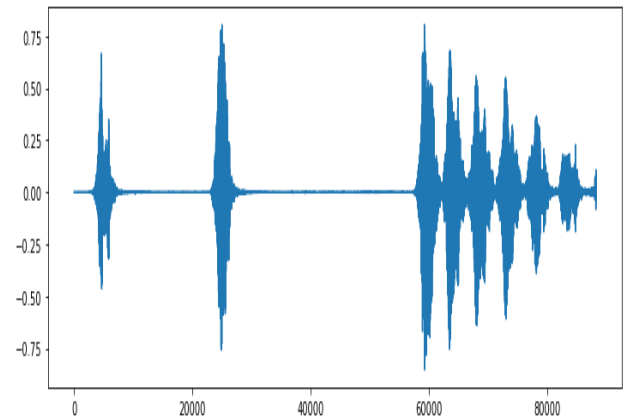
Figure 3

✓ Classification Report (Poor due to imbalanced dataset)
precision recall f1-score support.

	precision	recall	f1-score	support
0	0.09	0.09	0.09	199
1	0.06	0.05	0.05	83
2	0.12	0.14	0.13	200
3	0.11	0.09	0.10	190
4	0.13	0.13	0.13	206
5	0.09	0.10	0.10	201
6	0.11	0.08	0.09	90
7	0.12	0.12	0.12	206
8	0.13	0.14	0.14	189
9	0.13	0.14	0.13	183

accuracy			0.11	1747
macro avg	0.11	0.11	0.11	1747
weighted avg	0.11	0.11	0.11	1747

✓ Confusion Matrix of Random Forest Classifier (Poor due to imbalanced dataset)
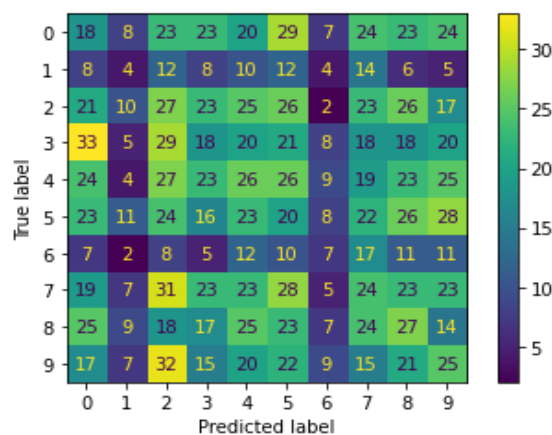

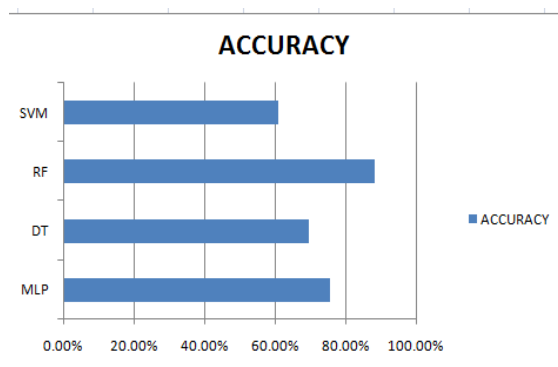Figure 4

✓ Summary of results:

Figure 5

From the above results table, we can see that our outcomes are very incredible. This is on the grounds that we did numerous analysis and experiments for extricating appropriate features and tracking down suitable methods. For each component, we attempt to get their actual implications and pick a suitable one for various utilizing purposes. For each neural network, we investigated it completely. From the above outcome table, we can see that our results are very extraordinary. This is on the grounds that we did numerous analysis and experiments for removing appropriate features and tracking down suitable methods. For each component, we attempt to get their actual implications and pick a suitable one for various utilizing purposes. For each neural network, we investigated it completely.

Conclusion: we have made project on sound order, exploratory information examination we done by utilizing urbansound8k dataset which utilized 10 unique classes of sound to recognize the class of sound by utilizing numeric identifier esteem, In this task, we have executed different AI techniques to acknowledge metropolitan sound grouping of ten classes. We have introduced a way to deal with sound grouping, which comprises of various highlights Firstly, we removed highlights from sound records in metropolitan sound 8K dataset and highlights,We have utilized four different neural organization models which are ANN, DT, RF and SVM. Every one of the models have been prepared and tried with unique UrbanSound8K and its increased dataset voice signs of 10 distinct classes were perused by utilizing the python library called Librosa which utilizes two sorts of signs one is mono and other is sound system. Moreover, we might want to continue with this in our future work where we have intended to utilize novel solo learning procedures that can be embraced to prepare, test the models, and really look at their exactness.

Future work and recommendations: With the assistance of sound order, while giving suggestions or offering types of assistance, applications in cell phones would not just base on the spot data be able to yet in addition base on encompassing sound data. Consequently, applications can offer

more precise support in view of general climate.

Assuming a gadget can give a portrayal of general climate in view of the sound order to the impaired, it will assist them with keeping away from certain risks and make their life more helpful.

References:

1. Li J, Wang Y, Zhu H, Zhang Y. What's That Sounds? Machine Learning for Urban Sound Classification.

2. Das JK, Ghosh A, Pal AK, Dutta S, Chakrabarty A. Urban sound classification using convolutional neural network and long short term memory based on multiple features. In2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS) 2020 Oct 21 (pp. 1-9). IEEE.

3. Jung DH, Kim NY, Moon SH, Jhin C, Kim HJ, Yang JS, Kim HS, Lee TS, Lee JY, Park SH. Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering. Animals. 2021 Feb;11(2):357.

4. Ruff ZJ, Lesmeister DB, Appel CL, Sullivan CM. Workflow and convolutional neural network for automated identification of animal sounds. Ecological Indicators. 2021 May 1;124:107419.

5. Nanni L, Costa YM, Aguiar RL, Mangolin RB, Brahnam S, Silla CN. Ensemble of convolutional neural networks to improve animal audio classification. EURASIP Journal on Audio, Speech, and Music Processing. 2020 Dec;2020(1):1-4.

6. Khamparia A, Gupta D, Nguyen NG, Khanna A, Pandey B, Tiwari P. Sound classification using convolutional neural network and tensor deep stacking network. IEEE Access. 2019 Jan 8;7:7717-27.

7. A Deep Neural Network for Audio Classification with a Classifier Attention Mechanism